Powerful outcome measures in MS

van den Elskamp, I.J.

2010

Chapter 4

General Discussion, Summary, and Future Perspectives

4.1 | General

The work presented in this thesis has centred on studying the statistical distribution and the statistical power of conventional and unconventional MRI outcome measures in clinical trials of MS. With the widespread use of approved therapies altering the practice of trials in MS, use of more sensitive and powerful outcome measures to maximize the ability of detecting treatment effects is becoming increasingly important. The first part of this thesis described two studies concerning the most frequently used MRI outcome measure in MS trials; the number and volume of enhancing lesions, and a study that applied a new, alternative measure for assessment of the inflammatory activity within MS patients: the number of T2w subtraction lesions. The second part focused on three outcome measures assessing neuroprotection and neurorepair: the number of persistent black holes, cerebral atrophy and lesional MTR. In this chapter, the main findings of the studies presented in this thesis are summarized and discussed. In addition, recommendations for future research will be presented.

4.2 | MEASURING INFLAMMATORY ACTIVITY

At present, the number of new Gd enhancing lesions is the most frequently used MRI outcome measure in MS clinical trials and, not surprisingly, most methodological progress has been made with this measure in the past decade. The first study addressing the issue of the required sample size for clinical trials using enhancing lesions as measure of choice, applied a non-parametric bootstrap resampling method with treatment effects simulated by Bernoulli trials. Based on a small and heterogeneous sample of MS patients it was shown that approximately 75 patients per arm were required to detect a 50% reduction in the number of active lesions in placebo controlled trials within 2 to 4 months respectively [1]. This methodology was adopted in three subsequent studies, and yielded comparable results for parallel grouped clinical trials based on small to medium sized cohorts of patients with homogeneous MS disease types [2,3,4].

While the applied methodology has the advantage of being independent from any assumption about the distribution of lesion counts, a major limitation of this approach is the resampling from small patient cohorts, leading to treated and untreated groups made up almost by the same subjects, thereby artificially decreasing

the between subject variability [5]. A more realistic approach is sampling patients from an infinite population, with a distribution of enhancing lesion counts described by parameters estimated from a genuine cohort of MS patients, rather than sampling from a small and fixed population.

Therefore, the statistical Negative Binomial (NB) model was proposed for describing the distribution of new enhancing lesion counts, showing an acceptable fit on a mixed cohort of RRMS and SPMS patients [5] and homogeneous cohorts of RRMS patients [6,7], and proved the previously applied non-parametric estimates to be underestimations compared to the NB distribution methodology, with an 50-100% increase of the number of required patients needed to redeem the same statistical power.

While the fits in these studies seemed adequate, it could not be ruled out that alternative distributions could be superior to the NB distribution. In **chapter 2.1** therefore, we aimed to explore the fit of several conceivable alternative statistical distributions for describing the number of enhancing lesions in both treated and untreated RRMS patients. We found that from the six statistical models addressed in our study the optimal and most constant distribution for modelling enhancing lesion counts remained the NB distribution, and that its fit was practically unaffected by the presence of a treatment effect, duration of follow up, and baseline MRI activity. However, only small differences in fit between the NB distribution and the alternative Poisson mixture models P-IG and P-LN were observed.

We concluded that the NB distribution is the distribution of choice for modelling new enhancing lesion counts; a finding with several implications [5]. First, the parametrization of the occurrence of newly enhancing lesions would allow actual estimation of the effect of treatments on lesional MRI activity instead of solely analyzing statistical significance. Second, by fitting multivariate regression models, the NB distribution provides a valuable tool for obtaining estimates of treatment effects, while adjusting for the effect of confounding variables. Thirdly, as aforementioned, it allows MS-trialists to perform more precise parametric sample size calculations.

Chapter 2.2 illustrates the applications of the NB described above. In this study, we compared monthly gadolinium-enhanced T1-weighted imaging with long-interval T2-weighted subtraction imaging within a 9 months placebo-controlled clinical trial of 116 RRMS patients, and evaluated the treatment efficacy by using both the nonparametric Mann-Whitney U test as well as a NB linear regression analysis. The trend in the results for both outcome measures was similar, with the Mann-Whitney

U test yielding unfavorable p-values, NB regression without baseline adjustment yielding smaller p-values and NB regression corrected for the number of lesions at baseline yielding the smallest, and clearly significant p-values. The gain in power, assuming an underlying NB distribution, resulting from the application of a parametric approach to the evaluation of treatment efficacy, thus shows the added value of studying the distribution of outcome measures in clinical trials. Whereas the results for T2w subtraction lesions are to be interpreted with care since the NB assumption for this parameter is based on a single dataset, the adequate fit of the NB distribution for T2w subtraction lesions was not unlikely with the majority of subtraction lesions overlapping with their Gd enhancing counterparts. With the assumption of both outcome measures to follow the NB distribution, the additional parametric sample size calculations in this study showed the distributional characteristics of both measures to differ to such an extent that the sample sizes required for T2w subtraction imaging to detect significant treatment effects were 22% to 34% lower compared to Gd enhanced imaging. Furthermore, the estimated sample sizes for Gd enhanced imaging were in line with previous results, a result providing additional validation of previous estimations [6].

In **Chapter 2.3** we take the first steps in exploring the distribution and statistical power of Gd enhancing volume as primary outcome measure in clinical trials. Enhancing lesion volume is a potentially more sensitive outcome measure for the detection of anti-inflammatory treatment effects compared to enhancing lesion counts, since total enhancing lesion volume accounts for both treatment effects on the number as well as on the size of lesions (which is known to diminish by treatment [8,9]). We found that enhancing lesion volume was a highly skewed variable with a large amount of discrete zero lesion volumes not amendable by transformation. Therefore, we applied a mixture of the binomial distribution and the Weibull distribution, where the binomial distribution modelled the proportion of inactive patients (fitting the applied datasets perfectly) and the Weibull distribution, shown to be the best fitting distribution from six conceivable alternatives, modelled the enhancing lesion volume of *active patients*. Next, we compared the required sample size for enhancing lesion counts with the required sample size for enhancing lesion volumes as primary outcome measure in clinical trials, by performing parametric sample size calculations using the methodology by Sormani *et al* [5], based on the NB distribution for enhancing lesion counts and the Binomial-Weibull mixture for enhancing lesion volumes. Although the comparison was complicated by the fact that the simulations for enhancing lesion

volumes needed to be performed with two treatment effects (one for the reduction in inactive patients and one for the reduction in lesion volume within active patients) and simulations for enhancing lesion counts with one (mean percentage reduction in lesion number), the estimated sample sizes still showed the order of magnitude of the sample sizes for enhancing lesion volumes to be considerably lower than the estimates for enhancing lesion counts, even when there was no increase in inactive patients and treatment effects are solely driven by a decrease in enhancing volume of active patients. This study thus indicates the potential gain of using enhancing lesion volume as primary outcome measure in clinical trials in terms of statistical power which, combined with the ongoing development of MRI automated segmentation methodologies, makes it not unlikely enhancing lesion volumes might become a more prominent measure of choice for future trials of immunomodulatory agents. At present, being merely the start of the methodological exploration of enhancing lesion volumes as outcome measure, it is premature to take definite decisions regarding its statistical distribution, and alternative modelling options such as the Tweedie distribution [10], which is capable of describing both zero lesion volumes as well as positive lesion volumes concurrently in a single simplified model, should be explored.

4.3 | MEASURING NEUROPROTECTION AND REPAIR

With multiple agents already available that target the inflammatory pathology of MS, the attention of research and clinical development in MS is turning increasingly to new therapeutic strategies for neuroprotection and neurorepair, thereby increasing the demand for pathologically specific and sensitive imaging outcome measures. The second part of this thesis we chose to explore the distribution and statistical power of outcome measures monitoring therapies with these specific properties. While the nomenclature in literature is not strict and terms are sometimes used interchangeably, in this thesis, neuroprotection refers to slowing of degeneration of neural tissue, and neurorepair to restoring nervous tissue integrity and function [11].

In **chapter 3.1** we consider the number of persistent black holes (PBHs) as primary outcome measure which, compared to BHs in general, specifically represent axonal loss and loss of myelin [12]. We proposed the NB distribution for modelling the cumulative number of PBHs at study end and showed its fit to surpass the fit of the Poisson distribution, the historical distribution of choice for modelling a counts

variable. Then, with a parametric simulation procedure based on the NB distribution, we showed that approximately 30 to 200 subjects per arm are required to detect significant treatment effects ranging from a 90% to 50% reduction of the number of PBHs respectively, and concluded that a potent drug is required to obtain sufficient power in clinical trials.

The choice for applying the NB distribution, as for the number of enhancing lesions and the number of T2w subtraction lesions is not a coincidence. First, the NB distribution is a well known solution for modelling overdispersed counts data, and lesional outcome measures in MS belong to this particular category. Second, all three measures are interrelated, and are likely to show similarities in distributional characteristics. Third, the choice for the NB distribution, as mentioned in **chapter 2.1** is also based on the principle of parsimony, which implies that the applied statistical model should not only closely fit the data, but should also be mathematically as straightforward as possible. Although also NB distributed, the estimated sample size proved different compared to the ones estimated for the number of enhancing lesions and number of T2w subtraction lesions.

An intriguing issue in our analysis is the role of the non-parametrical test in the simulation procedure on the eventual estimated sample size. At present, in the absence of a reasonable alternative, all calculations based on the methodology by Sormani et al. are completed with the application of the Wilcoxon test. **Chapter 2.2** already showed that, if the outcome measure in the population is assumed to be NB distributed, the application of a generalized linear model based on the NB distribution for analyzing treatment efficacy is more efficient in terms of study power. When applied in the current sample size calculations, a larger number of the 10.000 simulated trials would have reached statistical significance, and would have yielded more statistical power. Thus, the present sample size estimates, including those in previous studies for enhancing lesion number as outcome measure, might possibly be overestimations.

While the number of PBHs is a specific marker for neuronal damage, it comprises solely the damage occurring within lesions, *e.g.* focal pathology. Degeneration of nervous tissue in MS however is also known to occur diffusely in the surrounding non-lesional normal appearing brain tissue [13,14]. An MRI measure encompassing both focal as well as diffuse neurodegeneration in MS is the measurement of cerebral atrophy, for which various MRI techniques are available. As a global measure of all processes influencing the volume of the brain, a disadvantage of atrophy measures

is the lack of specificity for location and for tissue-specific processes, such as loss of myelin or axons and increase in glial content [15].

Recent studies assessed the required sample size to significantly detect treatment effects in clinical trials using cerebral atrophy as primary outcome measure. An initial study assessed a small cohort of RRMS patients using various measures of volumetric MR imaging for a follow-up of three years. Sample size calculations based on linear mixed model estimations showed SIENA to require 44 patients per arm to detect a 50% reduction in atrophy rate for a 2 year trial duration [16]. A second study supplemented this study by providing sample size estimates for patients in the SP stage of the disease, using MR scans obtained in an multi centre setting, and showed 32 patients to be required for the same efficacy and duration [17]. Both studies clearly suggested that SIENA (see introduction) should provide sufficient power to detect treatment effects. The focus of these studies however, was primarily on trials of long durations. Since phase II clinical trials in MS typically last only 6-12 months, we aimed to explore whether it was (statistically) feasible to perform short term clinical trials using cerebral atrophy as primary outcome measure. In **chapter 3.2** we assessed the rate of cerebral atrophy with SIENA in a large cohort of RRMS patients, followed for six consecutive months. Sample size calculations, based on the standard formula for normally distributed data, showed approximately 283 patients per arm are required in an unselected sampled population, and 185 patients in a population selected for a high T2 lesion load at baseline, to detect a 50% decrease in rate of atrophy. This shows that cerebral atrophy as primary outcome measure in short termed trials requires a potent drug to obtain sufficient power with feasible sample size.

Rate of cerebral atrophy is a normally distributed variable and thus, well known statistical methods and formulae based on this distribution apply. In previous studies linear mixed models were applied to take the dependency of longitudinal measurements into account, and considerable attention was given to the influence of trial duration on the required sample size [16,17]. In addition, a previous study [18] showed considerable gain in statistical power can be achieved for trials using cerebral atrophy as primary outcome measure by adding multiple scanning timepoints to a trial. By placing additional scans towards the start and end of the trial, the study showed reductions in total variance and hence reductions in trial size of 41% could be achieved in patients with Alzheimer disease, using the brain boundary shift integral method (BBSI) for determining the rate of cerebral atrophy. In particular, due to the within-subject variance contributing more to the overall variance at *short* intervals,

acquiring multiple scans has more impact in shorter studies. Although relatively smaller gains in power can be expected from adding time points for more precise measures such as SIENA, the effect on the required sample size should definitely be explored in future multi time point MS atrophy data. Unfortunately, this was not possible in the current dataset, since the intervening timepoints were not assessed for brainvolume changes.

Our finding of a significant rate of cerebral atrophy detected within a 6 months period (PBVC: -0.33%; SE 0.061; P<0.0001) added to the variable results of atrophy studies conducted over short intervals. At present, using various measures of brain atrophy assessment, two studies with a follow-up duration of 3 and 9 months did not detect a significant decrease in brain volume in RRMS patients, whereas two other studies of similar duration did detect a significant decrease [19,20,21,22]. The detectability of cerebral atrophy is thought to be related to the activity of the patients involved, with brain volume loss occurring at faster rates in active MS patients [22]. Since patients in the present study were selected for baseline activity (as is regular practice in MS clinical trials) the cohort can be regarded as fairly active. In addition, our study underlined the effect of selection for activity at baseline with the higher rates of atrophy in the active subgroups, and subsequent lower and more attainable sample sizes. Short interval brain atrophy assessment in the context of a clinical trial thus benefits from active patient cohorts in terms of detectability and study power. A potential drawback however, is the introduction of unwanted fluctuations in brain volume due to oedema and inflammation, thereby affecting the specificity of the outcome measure and reducing the power.

In a recent study, the rate of cerebral atrophy was determined with MRI data from the short interval oral Temsirolimus trial [23]. Temsirolimus demonstrated its immunosuppressive efficacy in the original study with a significant decrease in relapse rate and a reduction in the number of enhancing lesions on MRI in the 8mg group, but was also known to potentially exert a neuroprotective effect. With approximately 60 to 70 patients per arm, receiving either placebo or various doses of Temsirolimus for 9 months, a significant loss of brain volume (SIENA / PBVC: -0.33%) was found in the placebo group, and treatment with Temsirolimus significantly reduced the rate of brain atrophy in the 8mg group, with complete stabilization or even minor increase in brain volume (+0.14% (SE= 0.13%), P=0.0063) *vs.* placebo respectively), including dose-dependent effect on the rate of brain volume. This study illustrates the feasible

application of cerebral atrophy within a short period of time, given the beneficial effect of the treatment on cerebral atrophy is potent.

Chapter 3.3 aimed to determine the required sample size for clinical trials using the evolution of magnetisation transfer ratio (MTR) of newly formed lesions as primary outcome measure. MTR of cerebral tissue has greater pathological specificity for myelin content compared to conventional MRI, being high in undisrupted white matter and decreasing significantly within MS lesions [24]. The recovery of lesional MTR is therefore suggested as an outcome measure for clinical trials assessing the effect of therapies inducing remyelination, and thus realizing neurorepair.

In **chapter 3.3**, the MTR of newly formed lesions was assessed on 7 monthly MRIs within 32 patients from 5 centres. Even within this relatively small sample, we were able to determine the characteristic pattern of lesional MTR before and after enhancement, and modelled this evolution accordingly by means of multilevel models. The resulting estimates of the variance components were applied in power calculations to determine the required sample size for placebo-controlled, multicentre trials using lesional MTR recovery post-enhancement as primary outcome measure. These calculations showed that for a power of 80%, approximately 136 patients per trial (with an estimated mean number of 6 lesions per patient) are required to detect a 30% increase in lesional MTR post-enhancement compared to placebo, whereas 48 subjects are required to detect a 50% increase in lesional MTR compared to placebo, assuming a mild variation in the treatment-response between patients. Therefore, recovery of lesion MTR, appears to be a feasible outcome measure for future multicentre clinical trials measuring the effect of remyelinating agents.

An obvious difficulty in this chapter was to adequately model the evolution of lesional MTR, knowing MTR measurements to differ between centres [25], lesion evolution to be uniform within patients [26] and sequential measurements of single lesions to be highly dependent. Therefore, given the fact that lesional MTR data follows the normal distribution, we applied multilevel models to account for the dependency of the data on various measurement levels. From the various models that were applied on the available data, and compared by means of their maximum log likelihood, complex models including time as a factor did not improve the fit.
The chosen model describes the series of MTR measurements pre-enhancement and the series post enhancement by individual fixed slopes, and the difference between the first MTR measurement pre-enhancement and the last MTR measurement post-enhancement is allowed for using random slopes and random intercepts within lesions,

patients and centres. Because lesions were assessed for a maximum of seven months, and the total MTR evolution interval studied surrounding enhancement comprises 12 time points (time -6 to time +5), every single lesion yielded information of only a part of this interval. This explains the decreasing number of MTR measurements nearing time -6 and time +5. Mean MTR values for every timepoint pre and post enhancement were then calculated based on the available lesions, but corrected for their dependency on the various levels of measurement within the multilevel model. The relatively small sample of RRMS patients on which the analyses are based may have limited the modelling options described above. Larger datasets may result in complexer models to describe the data more accurately, and may subsequently influence the required sample size.

A recent, novel approach for monitoring the evolution of MTR changes has also shown promising results. Instead of the mean lesional MTR, the MTR of individual lesion voxels can be monitored, which potentially is a more sensitive method for picking up effects of experimental treatments [27].

4.4 | MODELLING MRI OUTCOME MEASURES: WHICH ASSUMPTIONS APPLY?

The sample sizes presented in this thesis were acquired using parametric calculation procedures. While the end results, *e.g.* the actual estimated sample sizes, are the most important findings of these analyses, one has to be aware of the applied assumptions when interpreting the estimates.

The first and most important assumption is obviously the statistical distribution used to describe the outcome measure. No matter how well a statistical distribution fit the data, one must realize that no mathematical formulae can replace any biological variable, and a measure of inaccuracy will always be present. The NB distribution plays an important role in this thesis for modelling the number of lesions as seen on MRI. In all calculations however, whether in those based on the number of enhancing lesions in previous studies or those described in this thesis for the number of T2w subtraction lesions or the number of PBHs, the dispersion parameter is assumed to remain constant and unaffected by treatment. It is conceivable that active treatments not only modify the mean number of lesions expected over a time period, but also the variability across patients. In **chapter 2.1**, we explored this issue by fitting the

NB distribution on data from patients effectively treated with an immunosuppressive compound [28]. This analysis showed that the dispersion parameter remained fairly constant and indicated that the use of a constant theta-parameter in parametric sample size estimations based on the NB distribution appears valid. Still, similar parameter estimates in other datasets should confirm this finding.

In **chapter 3.1** we assumed the evolution of newly formed lesions into a PBH within patients to be a random process. A previous study however showed that this evolution was a patient specific phenomenon, *e.g.* some patients are more prone for developing PBHs than others [26]. When this finding is taken into account, multilevel modelling might be a more adequate approach and subsequently could influence the estimated sample size. It was also assumed that new lesions of all lesion types had an equal chance of evolving into a PBH, whereas the analyses showed that larger lesions and for example ring enhancing lesions had a larger tendency to do so. Correction for size and lesion type thus needs to be explored as well.

Chapter 3.2 shows an example of an assumption which might prove unfeasible in practice and should be interpreted accordingly. The effect of a treatment is deemed to be effective from onset and constant over time, since it would be complex to model a treatment effect varying in time in terms of maximal efficacy. In practice however, a compound may take time to become maximally effective which thus decreases the detectable effect size when assessed beforehand, and subsequently increase the required sample size.

The calculations in **chapter 3.3** are based on several assumptions. First, we based our calculations on patients developing at least a certain mean number of lesions. Since in practice a proportion of patients will always prove inactive, sample sizes should be recalculated accordingly. Also, an important assumption was our definition of a 100% treatment effect. Although it has been shown that remyelinated lesions return a significantly lower MTR signal than the NAWM-MTR, we chose to define a 100% effect as an MTR value similar to the starting MTR value pre-enhancement, because of the absence of a valid reference. We clearly stated however, that the current sample size might be underestimated because of this assumption. In this chapter we also encountered two interaction terms in the calculations for which no estimate was available. Whereas we safely assumed the response to treatment would not vary among the participating centres, the difference in treatment response between patients could not be ignored, and sample size estimates were presented for interaction terms of varying size to assess their respective influence.

The above examples illustrate the importance of taking into account the assumptions applied when the estimated sample sizes of a study are considered, and shows the difficulties that need to be solved prior to a power analysis.

4.5 | MRI AS A SURROGATE FOR CLINICAL OUTCOME

The main focus of this thesis was to discuss the statistical modelling, power, and subsequent feasibility of using MRI measures as primary outcome in clinical trials. At present however, *clinical* outcomes from treatment trials of new agents for MS, such as relapse rate and measures of disability accumulation, are the only measures accepted by regulatory agencies. To enable MRI measures to act as surrogate outcomes for clinical trials in MS, MRI measures have to be rigorously validated according to strict criteria [29,30], thereby thoroughly confirming the relationship between both measures. No MRI outcome measure has currently been validated as an approved surrogate outcome in large phase III clinical trials. Unvalidated MRI surrogate measures are solely accepted in phase II clinical trials in MS providing proof of concept for potential new treatments. This thesis therefore, focuses primarily on phase II clinical trials, and largely passes over the issue of surrogacy. One should keep in mind that the feasibility of using an MRI measure as primary outcome not only depends on the statistical power of the measure, but also on the validity of substituting mandatory clinical measures.

Interesting developments in this regard have recently been shown for the relationship between traditional clinical outcome measures in MS and the number of enhancing lesions [31]. Whereas MRI markers are known to correlate poorly with clinical measures in MS, this study showed a novel approach by addressing the correspondence between the treatment effects observed for both outcomes instead of the direct relationship between the measures. The performed meta analysis showed that large effects on the number of enhancing lesions also had a large effect on the clinical outcome and vice versa, and convincingly demonstrated that MRI fulfils an important step in its validation as a surrogate marker.

These results hold promise for the potential validation of less conventional MRI outcomes as described in this thesis, and underline the importance of investigating the relationship between MRI and clinical measures, alongside the exploration of the statistical power of using MRI outcome measures in clinical trials.

FUTURE PERSPECTIVES

The results presented in this thesis substantiate the importance of the efficient use of MRI outcome measures of clinical trials. Based on the findings and the previously described relevance and limitations of the studies described, several recommendations for future research of statistical modelling and power of MRI outcome measures in clinical trials can be made.

The most general effort that should be made is to investigate the statistical distribution of MRI outcome measures in more, and preferably larger datasets. Throughout this thesis, decisions on the statistical modelling of the applied outcome measures were based on one or two datasets. Before a distribution can be routinely used in the design and analysis of MS clinical trials however, the adequacy of the distribution should be confirmed in multiple datasets. Although our datasets were of adequate size, compared to other fields of research involved with statistical modelling of outcomes, like economics and ecology [32] MS investigators have limited amounts of data at their disposal. Future modelling studies therefore, could benefit from large, multicentre databases like the one residing at the Sylvia Lawry centre for MS Research [33], and perform pooled analysis on placebo-treated patients from a large number of completed trials. Although not without limitations, being complex data with hidden sources of bias, variable MRI analysis techniques and potential insufficient durations of follow-up, modelling studies based on larger datasets will provide stronger evidence regarding the distributional nature of MRI measures.

Furthermore, additional studies using data from cohorts of effectively treated patients are needed to elucidate the possible distributional changes brought about by treatments, and their effect on the estimated sample sizes. Data with actual treatment effects are also needed to validate unknown variables now assumed to be of artificial size, such as the interaction terms in **chapter 3.3** describing the variation in treatment response between patients and between centres. In practice, this type of data may prove difficult to acquire since trial data is not easily available given the sensitivity of the information. Every endeavour however should be encouraged.

For the number of enhancing lesions, the adequacy of the NB distribution has been shown in multiple datasets in both **Chapter 2.1** as well as various previous studies. Furthermore, **Chapter 2.2** illustrated its application and added value in the analyses of clinical trials using new enhancing lesion counts as primary outcome measure. Future MS clinical trials should therefore routinely implement the NB

distribution in their design and data analysis when the number of enhancing lesions is applied as primary outcome.

An important aspect of future studies addressing the distribution of enhancing lesion volumes is the application of the Tweedie distribution [10] which, instead of considering lesional volumes as two separate circumstances (*e.g.* inactive patients and volumes in active patients) as proposed in **chapter 2.3**, models both zero volumes and volumes larger than zero concurrently in a single simplified model. Application of this distribution however is currently limited to specialized software and is not generally available in larger statistical packages. Future analyses should investigate whether this model indeed adequately fits enhancing lesion volume data, and is practically applicable. Another example for future studies exploring a different modelling approach is for the number of PBHs which, as mentioned in **chapter 3.1** might more adequately described using a multilevel model. Its result on the required sample size might prove interesting.

Another important goal for future studies is investigating the effect of alternative trial designs other than placebo controlled trials. With immunomodulatory therapies available that can reduce relapse rate in the short-term and decrease the accumulation of disability in the long term, it has become practically and ethically difficult to design and implement multiyear placebo controlled clinical trials. Future clinical trials will likely take the direction toward add-on trial designs, where new treatment in combination with an existing treatment is compared to the existing treatment alone [34]. Intriguing for trials assessing the efficacy of neuroprotective compounds for example, is the use of patients already receiving an anti-inflammatory compound, knowing the process of irreversible tissue loss to occur even in the absence of inflammation. Such designs will likely have consequences for the required sample size.

Future research should also address the distribution and statistical power of new MRI techniques. Studies regarding the required sample size for cerebral atrophy for example, have focussed on whole brain volume measurements, whereas in recent years, gray matter pathology has been recognized as an important feature in MS pathology and could prove to be very relevant to disability progression and cognitive decline [35]. While a recent study showed GMV measured with SIENAX to be a viable outcome in terms of study power, further explorations applying the preferred longitudinal measurement technique SIENA for example are anticipated [36].

REFERENCES

1. **Nauta JJ, Thompson AJ, Barkhof F, Miller DH.** Magnetic resonance imaging in monitoring the treatment of multiple sclerosis patients: statistical power of parallel-groups and crossover designs. *J Neurol Sci*. 1994;122(1): 6-14.
2. **Truyen L, Barkhof F, Tas M, Van Walderveen MA, Frequin ST, Hommes OR, Nauta JJ, Polman CH, Valk J.** Specific power calculations for magnetic resonance imaging (MRI) in monitoring active relapsing-remitting multiple sclerosis (MS): implications for phase II therapeutic trials. *Mult Scler.* 1997;2(6): 283-290.
3. **Tubridy N, Adèr HJ, Barkhof F, Thompson AJ, Miller DH.** Exploratory treatment trials in multiple sclerosis using MRI: sample size calculations for relapsing-remitting and secondary progressive subgroups using placebo controlled parallel groups. *J Neurol Neurosurg Psychiatry.* 1998;64(1): 50-55.
4. **Sormani MP, Molyneux PD, Gasperini C, Barkhof F, Yousry TA, Miller DH, Filippi M.** Statistical power of MRI monitored trials in multiple sclerosis: new data and comparison with previous results. *J Neurol Neurosurg Psychiatry*. 1999;66(4): 465-469.
5. **Sormani MP, Bruzzi P, Miller DH, Gasperini C, Barkhof F, Filippi M.** Modelling MRI enhancing lesion counts in multiple sclerosis using a negative binomial model: implications for clinical trials. *J Neurol Sci*. 1999; 1;163(1): 74-80.
6. **Sormani MP, Miller DH, Comi G, Barkhof F, Rovaris M, Bruzzi P, Filippi M.** Clinical trials of multiple sclerosis monitored with enhanced MRI: new sample size calculations based on large data sets. *J Neurol Neurosurg Psychiatry*. 2001; 70(4): 494-499.
7. **Sormani MP, Bruzzi P, Rovaris M, Barkhof F, Comi G, Miller DH, Cutter GR, Filipp M.** Modelling new enhancing MRI lesion counts in multiple sclerosis. *Mult Scler*. 2001; 7(5): 298-304.
8. **Di Rezze S, Gupta S, Durastanti V, Millefiorini E, Pozzilli C, Bagnato F.** An exploratory study on interferon beta dose effect in reducing size of enhancing lesions in multiple sclerosis. *Mult Scler 2007*; 13: 343-347.
9. **Gupta S, Solomon JM, Tasciyan TA, Cao MM, Stone RD, Ostuni JL, Ohayon JM, Muraro PA, Frank JA, Richert ND, McFarland HF, Bagnato F.** Interferon-beta-1b effects on re-enhancing lesions in patients with multiple sclerosis. *Mult Scler 2005*; 11: 658-668.
10. **Shono H.** Application of the Tweedie distribution to zero-catch data in CPUE analysis, *Fish. Res* 2008; 93; 154–162.
11. **Palace J.** Neuroprotection and repair. *J Neurol Sci*. 2008; 15; 265(1-2): 21-25.
12. **van Waesberghe JH, van Walderveen MA, Castelijns JA, Scheltens P, Lycklama a Nijeholt GJ, Polman CH et al.** Patterns of lesion development in multiple sclerosis: longitudinal observations with T1-weighted spin echo and magnetization transfer MR. AJNR *Am J Neuroradiol* 1998; 19: 675-683.
13. **De Stefano N, Iannucci G, Sormani MP, Guidi L, Bartolozzi ML, Comi G, Federico A, Filippi M.** MR correlates of cerebral atrophy in patients with multiple sclerosis. *J Neurol* 2002; 249: 1072-1077.

14. **Vrenken H, Geurts JJ, Knol DL, van Dijk LN, Dattola V, Jasperse B, van Schijndel RA, Polman CH, Castelijns JA, Barkhof F, Pouwels PJ.** Whole-brain T1 mapping in multiple sclerosis: global changes of normal-appearing gray and white matter. *Radiology* 2006; 240(3): 811-820.

15. **Anderson VM, Fox NC, Miller DH.** Magnetic resonance imaging measures of brain atrophy in multiple sclerosis. *J Magn Reson Imaging*. 2006; 23(5): 605-618.

16. **Anderson VM, Bartlett JW, Fox NC, Fisniku L, Miller DH.** Detecting treatment effects on brain atrophy in relapsing remitting multiple sclerosis: Sample size estimates. *J Neurol* 2007; 254:1588-1594.

17. **Altmann DR, Jasperse B, Barkhof F, Beckmann K, Filippi M, Kappos LD, Molyneux P, Polman CH, Pozzilli C, Thompson AJ, Wagner K, Yousry TA, Miller DH.** Sample sizes for brain atrophy outcomes in trials for secondary progressive multiple sclerosis. *Neurology* 2009; 72(7): 595-601.

18. **Schott JM, Frost C, Whitwell JL, et al.** Combining short interval MRI in Alzheimer's disease: implications for therapeutic trials. *J Neurol* 2006 ;253: 1147-1153.

19. **Zivadinov R, Bagnato F, Nasuelli D, Bastianello S, Bratina A, Locatelli L, Watts K, Finamore L, Grop A, Dwyer M, Catalan M, Clemenzi A, Millefiorini E, Bakshi R, Zorzon M.** Short-term brain atrophy changes in relapsing-remitting multiple sclerosis. *J Neurol Sci*. 2004; 223(2):185-93.

20. **Gasperini C, Paolillo A, Giugni E, Galgani S, Bagnato F, Mainero C, Onesti E, Bastianello S, Pozzilli C.** MRI brain volume changes in relapsing-remitting multiple sclerosis patients treated with interferon beta-1a. *Mult Scler.* 2002; 8(2): 119-123.

21. **Rovaris M, Comi G, Rocca MA, Wolinsky JS, Filippi M.** Short-term brain volume change in relapsing-remitting multiple sclerosis: effect of glatiramer acetate and implications. *Brain* 2001;124: 1803–1812.

22. **Hardmeier M, Wagenpfeil S, Freitag P, Fisher E, Rudlick RA, Kooijmans-Coutinho M, Clanet M, Radue EW, Kappos L, for the European rIFN B-1a in Relapsing MS Dose comparison trial study group.** Atrophy is detectable within a 3-month period in untreated patients with active relapsing remitting multiple sclerosis. *Arch Neurol*. 2003; 60:1736-1739.

23. **F. Barkhof, E. Fisher, I. van den Elskamp, M. Miller, M.M.S. Jasperse, R. Allen,R. Rudick, L. Kappos** The effect of oral temsirolimus on brain atrophy. *Mult Scler 2009 Ectrims suppl*. (abstract 767) In Press.

24. **Filippi M, Rocca MA, Martino G, Horsfield MA, Comi G.** Magnetization Transfer Changes in the normal appearing white matter precede the appearance of enhancing lesions in patients with multiple sclerosis. *Ann Neurol*. 1998; 43: 809-814.

25. **Tofts PS, Steens SC, Cercignani M, et al.** Sources of variation in multi-centre brain MTR histogram studies: body-coil transmission eliminates inter-centre differences. *MAGMA*. 2006 19(4): 209-222.

26. **Minneboo A, Uitdehaag BM, Ader HJ, Barkhof F, Polman CH, Castelijns JA.** Patterns of enhancing lesion evolution in multiple sclerosis are uniform within patients. *Neurology* 2005; 12; 65(1): 56-61.

27. **Chen JT, Collins DL, Atkins HL, Freedman MS, Arnold DL.** Magnetization transfer ratio evolution with demyelination and remyelination in multiple sclerosis lesions**.** *Ann Neurol*. 2008;63(2): 254-262.

28. **Kappos L, Barkhof F, Desmet A**. The effect of oral temsirolimus on new magnetic resonance imaging scan lesions , brain atrophy, and the number of relapses in multiple sclerosis: results from a randomised, controlled clinical trial. *J Neurol*. 2005;252 (suppl 2);46 (abstract 158).

29. **Prentice RL.** Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med*. 1989; 8(4): 431-40.

30. **Sormani MP, Bruzzi P, Comi G, Filippi M.** MRI metrics as surrogate markers for clinical relapse rate in relapsing-remitting MS patients. *Neurology. 2002*;12; 58(3): 417-421.

31. **Sormani MP, Bonzano L, Roccatagliata L, Cutter GR, Mancardi GL, Bruzzi P.** Magnetic resonance imaging as a potential surrogate for relapses in multiple sclerosis: a meta-analytic approach. *Ann Neurol*. 2009; 65(3): 268-275.

32. **Foster SD, Dunstan PK.** The Analysis of Biodiversity Using Rank Abundance Distributions. *Biometrics*. 2009; 4; In Press.

33. **Schach S, Scholz M, Wolinsky JS, Kappos L.** Pooled historical MRI data as a basis for research in multiple sclerosis--a statistical evaluation. *Mult Scler.* 2007;13(4): 509-516.

34. **Polman CH, Reingold SC, Barkhof F, Calabresi PA, Clanet M, Cohen JA, Cutter GR, Freedman MS, Kappos L, Lublin FD, McFarland HF, Metz LM, Miller AE, Montalban X, O'Connor PW, Panitch H, Richert JR, Petkau J, Schwid SR, Sormani MP, Thompson AJ, Weinshenker BG, Wolinsky JS.** Ethics of placebo-controlled clinical trials in multiple sclerosis: a reassessment. *Neurology*. 2008; 25; 70(13):1134-40.

35. **Geurts JJ, Bö L, Pouwels PJ, Castelijns JA, Polman CH, Barkhof F.** Cortical lesions in multiple sclerosis: combined postmortem MR imaging and histopathology. *AJNR Am J Neuroradiol*. 2005; 26(3): 572-577.

36. **Healy BC, Valsasina P, Filippi M, Bakshi R.** Sample size requirements for treatment effects using gray matter, white matter and whole brain volume in relapsing-remitting multiple sclerosis.*J Neurol Neurosurg Psychiatry*. 2009 Epub; 80(11): 1218-23.